scientific reports

OPEN

Fed-CL- an atrial fibrillation prediction system using ECG signals employing federated learning mechanism

Fayez Saud Alreshidi[1], Mohammad Alsaffar[2], Rajeswari Chengoden[3]✉ & Naif Khalaf Alshammari[4]

Deep learning has shown great promise in predicting Atrial Fibrillation using ECG signals and other vital signs. However, a major hurdle lies in the privacy concerns surrounding these datasets, which often contain sensitive patient information. Balancing accurate AFib prediction with robust user privacy remains a critical challenge to address. We suggest Federated Learning , a privacy-preserving machine learning technique, to address this privacy barrier. Our approach makes use of FL by presenting Fed-CL, a advanced method that combines Long Short-Term Memory networks and Convolutional Neural Networks to accurately predict AFib. In addition, the article explores the importance of analysing mean heart rate variability to differentiate between healthy and abnormal heart rhythms. This combined approach within the proposed system aims to equip healthcare professionals with timely alerts and valuable insights. Ultimately, the goal is to facilitate early detection of AFib risk and enable preventive care for susceptible individuals.

Atrial Fibrillation (AFib) stands as a prevalent and concerning cardiac arrhythmia affecting individuals globally[1–3]. The irregular and often rapid heartbeat associated with AFib poses substantial risks to cardiovascular health, making it a significant contributor to morbidity and mortality[4]. With the global population aging and the rise of predisposing factors such as hypertension and diabetes, the prevalence of AFib continues to escalate[5]. The implications of AFib in cardiovascular health cannot be overstated, as it is closely linked to severe complications, including stroke, heart failure, and an increased risk of mortality[6]. The irregular heartbeat characteristic of AFib disrupts the normal blood flow, potentially leading to the formation of blood clots in the heart chambers[7,8]. When these clots migrate to the brain, they can result in a stroke, cementing AFib as a leading cause of this debilitating condition. Traditional methods of AFib detection, while valuable, often face limitations in terms of accuracy and efficiency[9–11].

The need for precise and efficient detection methods is critical, considering the escalating morbidity and mortality rates associated with AFib. Early identification of AFib is paramount for timely clinical intervention and the implementation of appropriate therapeutic strategies. As the world enters an era marked by technological advancements, there exists a promising avenue for transforming AFib detection. ML and DL techniques have emerged as powerful tools with the potential to revolutionize cardiovascular diagnostics[12–14]. Harnessing the capabilities of these computational approaches holds the promise of not only enhancing the accuracy of AFib detection but also streamlining clinical diagnoses and interventions[15]. Although ECG is the most widely used and successful method, other conventional techniques can also be used to predict AFib, however, they might not be as accurate. Conventional techniques include, physical examination, medical history, pulse palpitation, and blood examination. With regard to sensitive data (such as financial or medical details), Federated Learning (FL) permits model training on decentralized datasets without the need for direct data sharing. This allows for collaborative learning without compromising user privacy. FL makes it easier to train a model on fragmented datasets when there are several owners of the data who are hesitant to share it in one place. This is helpful in contexts such as dispersed device learning (e.g., smartphone training) and collaborative research. Large datasets must frequently

[1]Department of Family and Community Medicine, College of Medicine, University of Ha'il, Ha'il, Saudi Arabia. [2]Department of Computer Science and Software Engineering, College of Computer Science and Engineering, University of Ha'il, 81481 Ha'il, Saudi Arabia. [3]School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India. [4]Mechanical Engineering Department, Engineering College, University of Ha'il, 8148 Ha'il, Saudi Arabia. ✉email: rajeswari.c@vit.ac.in

be sent to a central server via traditional methods, which can be computationally and bandwidth-intensive. FL is appropriate for contexts with limited resources since it maintains the data on each devices and only needs a minimal amount of connectivity to update the model. FL models may be able to capture a more varied range of attributes through training on a larger variety of data that is dispersed across several devices or places, improving their generalizability for practical uses. The suggested FL-based model for atrial fibrillation (AFib) prediction offers potential advantages in terms of accuracy and privacy while addressing the drawbacks of conventional techniques. Concerns about privacy arise because current AFib prediction methods frequently rely on centralized data collecting. Sensitive medical data is kept on the user's device while this model combines FL to train the model on dispersed data at local devices. Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) are combined in this model. While LSTMs are better at capturing the sequential dependencies in the data and may result in more robust feature learning for AFib identification, CNNs are better at extracting local features from ECG signals. The model recognizes that, for variable-length and imbalanced ECG datasets, data pre-processing is crucial. The data is prepared for efficient model training through the use of strategies like data transformation, standardization, and balancing.

This article embarks on an exploration of novel methodologies to address the current challenges in AFib detection, aiming to contribute to the advancement of cardiovascular healthcare practices. Motivated by the recognition of persistent challenges and limitations in traditional methods, the study delves into Federated Learning (FL) framework, offering an opportunity for improved privacy, accuracy and efficiency in AFib diagnosis. The traditional models have various limitations with respect to data privacy and security, data diversity and generalization, model personalization, scalability, efficiency, communication cost and latency. FL helps in preserving data privacy and security by keeping the data private at the client devices itself. FL helps in enhancing the model's generalization by training the model from diverse data sources. Collaboration among various organizations can also be made possible with the help of FL mechanisms. FL enables the development of personalized models that can be fine-tuned on local data, thereby improving the accuracy of the model. Since the computational model is distributed across different devices or servers by making the training process more scalable and efficient. The research sets clear objectives, to evaluate the efficacy of these computational techniques, assess the impact of feature extraction methods, investigate the role of data balancing techniques, and contribute valuable insights to the field of cardiovascular health and medical diagnostics. The scope of the research encompasses a detailed exploration of FL models for AFib detection, utilizing diverse types of medical data, primarily focusing on ECG signals. While the study ambitiously covers a wide array of models, methodologies, and data types, it acknowledges certain limitations, operating within the constraints of available datasets and computational resources. Motivated by persistent challenges in traditional methods, this research explores a threshold-based approach to detect anomalies in ECG signals, serving as an indicator for irregular cardiac rhythms (Pander, 2022). The motivation lies in addressing key issues of dataset imbalance, feature extraction, and preserving privacy of the data through innovative mechanisms. Particularly, the introduction of a hybrid CNN-LSTM based FL model (Fed-CL) demonstrates superior model evaluation metrics such as accuracy, recall, precision and F1 Score in predicting AFib. Figure 1 depicts the detection of AFib using ECG signals with the help of such techniques.

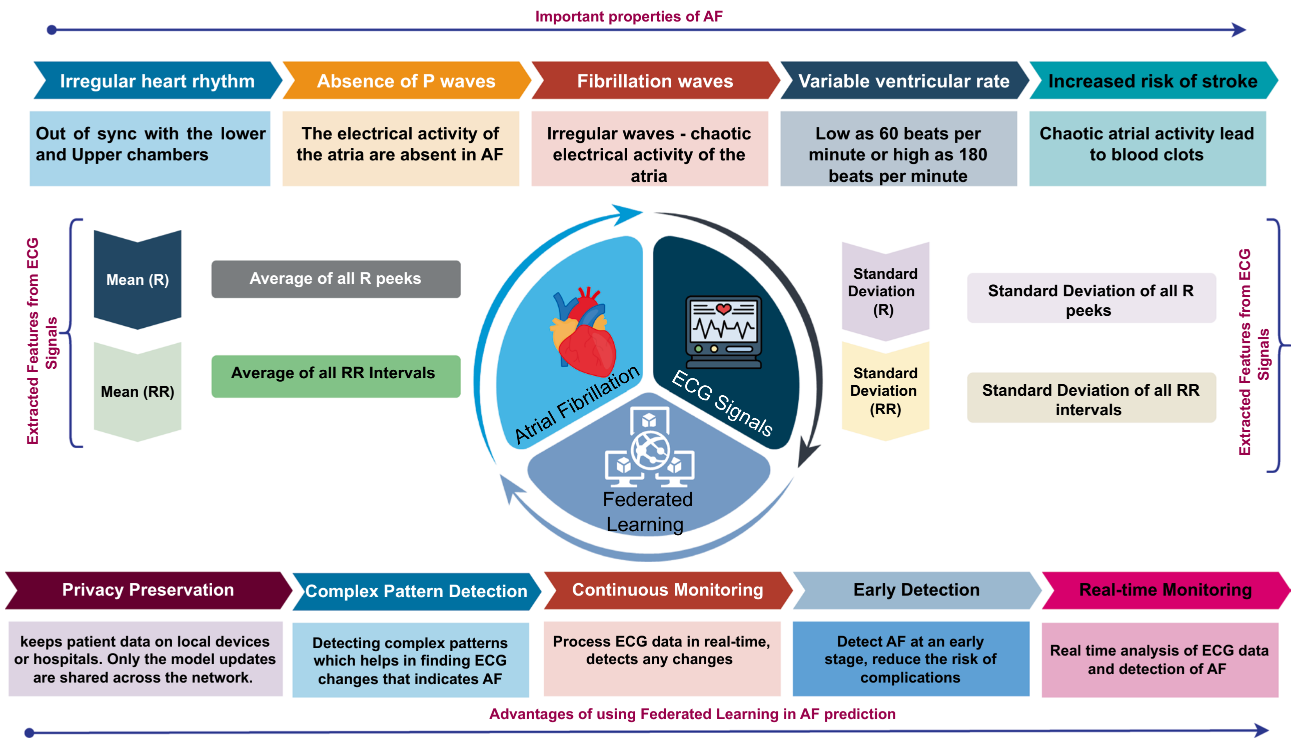


Figure 1. Detection of Atrial Fibrillation using ECG signals with the help of Federated Learning.

The primary objectives encompass evaluating the efficacy of the proposed prediction system, emphasizing the significance of mean HRV in distinguishing normal and abnormal heart rhythms. Hence, the study aims to contribute to early detection and preventive care for individuals at risk of AFib, providing valuable insights and timely warnings to healthcare professionals.

While existing studies have made valuable contributions to AFib detection, several limitations persist. Common challenges include issues related to model accuracy, dataset imbalances, impact of signal noise on diagnostic outcomes, data privacy, communication cost and latency. Many conventional methods lack adaptability to early and subtle manifestations of AFib, leading to delayed diagnoses and hindering effective interventions. The present study addresses the limitations of current research by introducing a AFib prediction system with distinctive contributions. Firstly, it systematically evaluates the efficacy of ML, DL and FL techniques, providing insights into the strengths and weaknesses of various models in identifying AFib patterns. Secondly, the study explores the impact of feature extraction techniques on model performance, offering a nuanced understanding of Pre-Processing methods. Furthermore, the research investigates the influence of Synthetic Minority Over-sampling Technique (SMOTE) on model performance, contributing to the understanding of handling imbalanced datasets. The introduction of a FL model represents a significant advancement, demonstrating superior accuracy, precision and recall in AFib prediction.

Thus, this study contributes to the refinement of diagnostic methodologies for AFib, overcoming current limitations, and paving the way for more accurate, efficient, and timely detection. The proposed prediction system aims to bridge existing gaps, providing healthcare professionals with a reliable tool to enhance patient care and preventive measures for individuals at risk of AFib.

The subsequent sections of this paper are structured to provide an exploration of the research arena. Section 2 discusses existing works related to the studies on prediction of AFib, establishing the contextual foundation. Section 3 introduces the materials and methods used in the proposed study. Section 4 introduces the experimentation set up of the study. The section also outlines the results and their interpretation, serving as the empirical anchor. Finally, section 5 concludes the paper, offering a synthesis of findings and outlining future research directions.

Related work

This section presents an extensive review of current research on AFib detection, predominantly concentrating on methods utilizing ECG signals. The literature is systematically categorized based on the types of algorithms employed, distinguishing between traditional ML and DL techniques.

The authors in[16] summarized the various ML models that can be employed in the detection and management of atrial fibrillation, along with the possible future areas of development. Another review was conducted by Siontis, K. C et. al[17] on the contributions from various DL and other ML techniques on the treatment of atrial fibrillation (AF). The authors in[18] proposed a ML technique to classify Intracardiac Electrical Patterns during Atrial Fibrillation. In a study evaluating ML approaches[19], researchers delved into predicting new-onset AFib using two distinct models - CNN and eXtreme Gradient Boosting. Their investigation involved raw ECG signals and signal-extracted features, with CNN showcasing superior performance over XGB and Logistic Regression. The study emphasized the crucial consideration of the approach based on the available data quantity. A meta-analysis of ML and DL models[20] focused on AFib detection, classification, and prediction using 12-lead ECG recordings. CNNs demonstrated excellence in AFib detection, while Support Vector Machines (SVMs) outperformed in classification. The study reported notably high pooled sensitivity and specificity for both detection and classification tasks. In the pursuit of optimized AFib detection with CNN[21], a novel approach was employed. This study not only utilized CNNs on ECG signals but also implemented noise reduction with a Butterworth filter and applied Short-term Fourier transform for ECG spectrogram images. The method yielded impressive performance metrics on the MIT-BIH AFib database. Another noteworthy study proposed a novel deep-learning approach[22] for automated AFib detection from single-lead ECG signals. The methodology involved Continuous Wavelet Transform (CWT) for time-frequency feature extraction and a specialized CNN with ResNet for effective training. This model outperformed traditional DL models, specifically addressing challenges associated with imbalanced data.

The study focusing on a CAELSTM combination model[23] aimed at predicting AFib, particularly paroxysmal AFib (PAF). The combination of unsupervised Convolutional Autoencoder (CAE) and supervised LSTM models showed promising accuracy, indicating potential usefulness in scenarios with limited labeled data. In a unique approach combining wavelet transform-based filtering and neural networkbased classification[24], the study proposed an effective methodology for early AFib detection. The achieved high accuracy and ROC values were further enhanced by 5-fold cross-validation.

Exploring Photoplethysmogram (PPG) signals as an alternative for AFib detection[24], another study employed a hybrid 1D CNN and Bidirectional LSTM. This innovative approach showcased high accuracy and effectiveness in identifying and classifying AFib cases. The automated detection of heartbeat abnormalities[25] involved the development of two ML models - an 8-layer 1D CNN and a combination of 1D CNN and LSTM. These models aimed at efficient detection and classification of heartbeat abnormalities, showcasing high accuracy, precision, and F1 score, potentially reducing diagnosis time. Summarizing studies utilizing DNNs for predicting heart disease, the review highlighted the enhanced accuracy achieved through advanced computational approaches. Various DNN architectures and their applications in cardiac disease prediction were discussed, emphasizing their effectiveness in handling large datasets and reducing computational load. In a broader perspective, the discussion on various applications of DNN models showcased the versatility of these models in cardiac disease prediction. Their capability to handle diverse datasets and reduce computational load was emphasized, making them a valuable asset in disease prediction.

FL enables ML models to be trained across multiple decentralized devices that store local data samples without exchanging data. In healthcare systems, patient data sensitivity and privacy are important. Hospitals and research institutions can collaborate to improve AFib predictive models without disclosing patient information. FL employs a decentralised AI model training Approach which allows collaborative learning without data exchange. The work in[26] presents an approach for using FL to train AI models for arrhythmia detection with 12-lead ECG signals attained from various sources. The primary goal is to create a privacy-preserving methodology for training AI models on high-definition ECG data from 12-lead sensor sets without sharing sensitive medical data, thus resolving privacy concerns associated with medical data usage. The experiment aims to examine the performance of AI models trained with the FL methodology on the PhysioNet in Cardiology Challenge 2020 dataset, which contains high-definition 12-lead ECG recordings from 43,059 patients from six unique geographical centers. The prediction methods use ML techniques using DNN and LSTM models, as well as a robust data preprocessing process that includes feature engineering, selection, and data balancing techniques. The research's findings show that detecting cardiovascular abnormalities enhances prediction rates compared to other existing models. However, we discovered that the authors did not explain how the data augmentation was performed in the classification of the 12 ECG arrhythmias. In[27], the authors developed a framework which integrates FL and the Modified Artificial Bee Colony optimization algorithm in order to enhance heart disease prediction accuracy while also addressing data privacy risks in healthcare. The main goal of the study is to improve heart disease prediction accuracy, training time, and communication effectiveness while maintaining data privacy. The experiment was carried out on a real-world heart disease dataset from the UCI Cleveland database, which included 303 records and 76 attributes and focused on different risk factors for heart disease. The framework had a prediction accuracy of 92.89% on the heart disease dataset. Additional metrics such as precision 94.2%, sensitivity 96.6%, specificity 81.8%, classification error 11.8% were obtained. The future work can be extended in addressing scalability concerning the number of Internet of Medical Things client sites and the effect of the learning rate on the overall model. Future research can concentrate on using feature selection algorithms and optimization methods to address a wider variety of health-related diseases, with special focus on privacy concerns in healthcare systems.

In[28], the authors proposed a hybrid classifier-based FL framework for cardiovascular disease prediction, with a focus on addressing privacy concerns, training latency, and communication costs by combining Modified Artificial Bee Colony Optimization with Support Vector Machine for optimal feature selection and classification. The main objective of the work is to propose a hybrid FL framework that improves heart disease prediction accuracy while also addressing data privacy concerns for health care providers. The framework aims to reduce training time and communication costs while improving prediction accuracy. The experiment used a combined dataset from Cleveland, Stalog, Hungary, Long Beach, and Switzerland, which included 11 common features for cardiovascular disease prediction. The dataset contains a variety of parameters, including blood pressure, cholesterol serum, and chest pain. The proposed hybrid technique outperformed FL techniques such as FedAvg-SVM and FedMA-SVM, achieving a maximum accuracy of 93.8% after 2408 communication rounds. The authors of[29] developed a weighted FL technique for ECG arrhythmia classification. This method allows for the heterogeneity of data distribution across several clients in FL environment, dynamically adjusting each client's weight based on its contribution to global model enhancement. The main purpose of the study is to deal with privacy concerns and improve accuracy in ECG arrhythmia detection. The experiments were carried out using publicly available ECG datasets, namely the MIT-BIH arrhythmia dataset, which contains 109446 samples with 5 classes and a sample rate of 125Hz. The prediction method employs a DNN that was trained using the weighted FL technique. This method dynamically adjusts the weight of each client's data based on how it contributes to overall model improvement. The features extracted from the ECG signals contain 13 input features, with the last column providing the output feature, which indicates whether the patient has a cardiovascular problem. On the client side, the accuracy percentage is 93%, with 98% sensitivity, 82% specificity, a 70% miss classification rate, and 60% precision. On the server side, the FL approach had 98% accuracy, 99% sensitivity, and 91% specificity. From the related works studied, different methods based on conventional ML, DL, and FL are utilised. Among these methods, decision-making processes can be better understood by using ML models, which are frequently simpler to comprehend and interpret. Compared to DL models, training ML models usually takes less computing power. Unlike DL models, ML model can function well even with a small amount of data. However, ML requires manual feature engineering, which can take a lot of time and requires domain-specific knowledge and there is possibility of overfitting on smaller datasets. DL works well on huge datasets specifically issues involving sensor readings and other high-dimensional data and can automatically extract intricate characteristics from unprocessed data, overcoming the necessity for human feature engineering. However, understanding the reasoning behind forecasts might be tough due to the complexity and difficulty of some models. In case of FL, training the models on dispersed data at local devices, the sensitive data is kept secret. It has the ability to use data from several devices without the need for centralized data. But for devices with limited resources or bandwidth networks, frequent communication between devices and the central server may be inefficient. Non-uniform data distribution among devices can cause problems with model convergence. Because training is a distributed process, it may converge more slowly than centralized approaches.

The comprehensive overview presented in Table 1 summarized different AFib detection methods, highlighting key details such as data sources, prediction methods, feature extraction techniques, and accuracy metrics. This section concludes by providing a comparative insight into the advancements made in ML, DL, and FL techniques for AFib detection, setting the stage for subsequent discussions on the problem statement, research objectives, and chosen methodology.

Ref. no.	Data source	Prediction method	Feature extracted	Accuracy	Additional metrics
[19]	ECG signals	CNN, XGB	Heart rate variability (HRV), P-wave features, T-wave features, ST segment features, QRS complex features	CNN: 98%	For CNN, calibration index changed from 0.014 [0.01, 0.018] to 0.17 [0.16, 0.19] after imbalance corrections
[21]	MIT-BIH AF database	CNN	NA	Accuracy: 99.4%	Sensitivity: 99.72%, Specificity: 98.95%
[22]	Real-world databases ECG	Specialized CNN with ResNet	Time-Frequency Features via CWT	Not Specified	Not Specified
[23]	20 segments of PAF and 20 normal heart rates	CAE-LSTM	RR interval time series	Accuracy: 93.56%	RMSE: 0.004, F1 Score: 0.9345
[24]	Publicly available ECG datasets	Wavelet transform, Neural Network	Filtered ECG signal	Dataset 1: 96%, Dataset 2: 86%	ROC: 0.95, 0.84
[26]	PhysioNet in Cardiology Challenge 2020 dataset, which comprises high-definition 12 lead ECG recordings from 43,059 patients.	DNN and LSTM models, incorporating a robust data preprocessing pipeline with feature engineering, selection, and data balancing techniques.	The cardiac rhythm of consecutive heartbeats, resulting in about 650 features	Not Specified	Not Specified
[27]	UCI Cleveland heart disease dataset	Combines FL with a Modified Artificial Bee Colony (M-ABC) optimization algorithm for feature selection and classification	Features related to heart disease risk, such as age, sex, blood pressure, serum cholesterol, hereditary factors, and more, were optimized using the M-ABC algorithm.	Accuracy: 92.89%	Precision (94.2%), Sensitivity (96.6%), Specificity (81.8%), Classification Error (11.8%)
[28]	Dataset from Cleveland, Stalog, Hungary, Long Beach, and Switzerland, consisting of 11 common features for cardiovascular disease prediction	Integrating Modified Artificial Bee Colony Optimization (M-ABC) for feature selection and SVM for classification.	Blood pressure, cholesterol serum, chest pain	Accuracy of 93.8% within 2408 rounds of communication	Precision (94.2%), sensitivity (96.6%), specificity (81.8%), classification error (11.9%), and F-measure (90.1%).
[29]	MIT-BIH arrhythmia dataset, which contains 109446 samples with 5 classes	Employs a DNN that was trained using the weighted FL technique	The features extracted from the ECG signals contain 13 input features	Accuracy percentage is 93%,	98% sensitivity, 82% specificity, a 70% miss classification rate, and 60% precision.

Table 1. Comparison of Different Methods for AFib Detection.

Research gap analysis

While the existing body of literature on AFib detection through ECG signals showcases substantial advancements, there are discernible research gaps that warrant attention. One notable gap revolves around the need for further exploration into the robustness and generalizability of the proposed models. Several studies have demonstrated commendable accuracy and effectiveness in controlled environments, utilizing specific datasets. However, a comprehensive analysis of model performance across diverse datasets and real-world scenarios is essential for evaluating their practical applicability. Moreover, the interpretability of ML models remains an area requiring more indepth investigation. Many existing studies rely on complex DL models, often perceived as "black boxes," hindering their adoption in clinical settings. The black box nature of DL models, are not transparent and difficult to understand. False positives, where the model incorrectly predicts AFib when it's absent, can lead to unnecessary medical procedures and patient anxiety. With limited interpretability, clinicians might not be able to explain or justify the model's predictions, raising concerns about accountability and potential biases within the model. A deeper understanding of the interpretability of these models and their diagnostic value is crucial for fostering trust among healthcare practitioners. Additionally, there is a noticeable gap in research addressing the impact of demographic and clinical variations on model performance. Exploring the efficacy of AFib detection models across different demographic groups and varying clinical conditions is crucial for ensuring the inclusivity and reliability of these models in diverse patient populations. However, FL faces some challenges, such as data heterogeneity, which results in diverse data distributions across multiple devices may prevent the building of a globally effective model. This heterogeneity can have an impact on the model's ability to generalize across the wide range of patient data. FL models may not generalize well to other populations with varied demographics, health conditions, or device kinds if they were trained on data from a particular population (such as a particular hospital). For patients not included in the training data distribution, this may result in imprecise predictions. Bridging these research gaps will contribute significantly to advancing the field of AFib detection and fostering the integration of FL models into clinical practice.

Materials and methods

This section delves into the fundamental concepts of Federated Learning within the context of predicting AFib from ECG signals. Federated Learning is employed for atrial fibrillation prediction majorly due to two reasons. Firstly, as the data for disease prediction model is large in quantity, it would be difficult to collect, process and analyse the same. Secondly, medical data is so sensitive and improper handling may lead to various privacy attacks and possible health hazards. Federated Learning comes to the rescue in such contexts by restraining the attack surface, if any, only to edge devices, rather than the devices and the cloud.

The model CNN-LSTM

A hybrid CNN-LSTM model named, MCNN-LSTM (Model CNN-LSTM) is employed here for building the global model. LSTM, a variant of RNNs, addresses the limitation of long-term dependency in RNNs by introducing a specialized memory cell. The memory cell, governed by input, forget, and output gates, enables the network to selectively retain or discard information, allowing for the processing of timeseries data effectively. The equations 1 to 6 demonstrate the intricate operations performed by the LSTM unit, showcasing its ability to maintain long-term dependencies.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_t) \tag{2}$$

$$\widetilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{3}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \odot \tanh\left(C_t\right) \tag{6}$$

Any LSTM unit usually consists of three basic steps: To determine what is to be forgotten from the cell state, what is to be kept in the cell state and what is to be given as output from the current cell state. In equation 1, it determines the information that is to be discarded from the previous cell state, f_t, σ is the sigmoid activation function, W_f is the weight matrix associated with the forget gate, x_t is the input at time t, h_{t-1} is the hidden state from the previous time step and b_f is the bias term. In equation 2, it represents the input gate at time t, which controls the information to be stored in the memory cell, σ is the sigmoid activation function, x_t is the input at time t and h_{t-1} is the hidden state from the previous time step, W_i is the weight matrix associated with the input gate and b_i is the bias term. Additionally, a"tanh" layer is incorporated that creates a set of new candidate values that is to be added to the current state. This is represented in equation 3 and 4. Employing computational optimizations like matrix factorization methods reduces the complexity and increases the efficiency of LSTM computations. Assessing these modifications enables a balanced approach for enhancing the performance and increases the computational efficiency, especially in a FL environment which are having limited computational resources and bandwidth.

Similarly, equations 5 and 6 helps in deciding what is to be given as output from the current cell state. This is determined by a sigmoid layer that decides the parts of the cell state that is to be given as the output. This also constitutes a"tanh" layer that outputs the parts, that represents the new cell state. W_o is the weight matrix associated with the output gate and b_o is the bias term. Equation 6 includes element-wise operations between the output gate and the tanh of the cell state. Pruning is considered in LSTM computation pipeline, that removes non-critical parts of the model weights, minimize computational load and improves the inference time while maintaining model performance.

Therefore, all the equations from 1 to 6, collectively describe the intricate computations within an LSTM unit, enabling the network to effectively capture and utilize information over extended time periods.

$$\hat{y}(x) = \max(\text{pool}(x)) \tag{7}$$

However, in equation 7, $\hat{y}(x)$ represents the max-pooling operation applied to the input x, wherein, Max pooling is a downsampling operation commonly used in CNNs. It involves dividing the input into non-overlapping regions and taking the maximum value from each region, reducing the spatial dimensions of the input.

$$``weight'' = ``weight'' - learningrate \times error \times input \tag{8}$$

Further, the equation 8 represents the weight update rule commonly used in the backpropagation algorithm during the training of neural networks, including CNNs. "weight" denotes a specific weight parameter in the network. learning rate is a hyperparameter that scales the size of the weight updates, influencing the convergence speed of the training process. error is the difference between the predicted output and the actual (ground truth) output. input is the input value that caused the error. The update is performed iteratively, and this process continues until the network converges, meaning it minimizes the error on the training data.

Additionally, the role of CNN in feature extraction is discussed, where convolutional layers use filters to highlight specific features in the input image. The convolutional layer is followed by a pooling layer, which reduces the size of feature maps, and a fully connected layer for classification. Backpropagation is employed for training, updating weights based on the error between predicted and actual values, ensuring the network convergence[3031].

Privacy measures

Privacy plays a significant role in the scenarios where personal data is involved. When it comes to dealing with medical data, the need to preserve the data privacy is paramount. The general definition for privacy is as follows: Definition 1 (Privacy): Consider F as the flow of the message M from a sender S to a receiver R through a channel C. Then, privacy refers to S's capacity to set an upper limit on the information contained in M, and of any computation on M, regardless of the R's prior knowledge.

Data anonymization techniques are usually employed to preserve the privacy of data that is shared to the third parties. Data anonymization can be defined as follows:

Definition 2 (Data Anonymization): Consider a dataset D with a set of sensitive attributes, $S(D)$ be the dataset after suppression, $G(.)$ be the generalization function and $P(.)$ be the pseudonymization function, then data anonymization can be defined as: $D \rightarrow P(G(S(D)))$.

Usually in a healthcare setting, the anonymized data only will be shared to the central server for training the model. Since the data is anonymized, the accuracy of the results will comparatively be low when compared to the data without anonymization. Model auditing and privacy preserving model training frameworks acts as a significant role in protecting against inference attacks and prevents the models from leaking sensitive information. Model auditing aids in recognizing how a model takes decisions and ensuring that these decisions do not unintentionally disclose sensitive information. Model auditing ensures that models do not unintentionally leak information due to errors in predictions. Privacy-preserving model training frameworks employ noise in the outputs to reduce the risk of inference attacks and guarantee the model does not leak sensitive information even when queried with complex questions.

Various anonymization techniques such as k-anonymity[32], l-diversity[33], t-closeness[34] and differential privacy[35] have widely been used for anonymizing the data for privacy preservation. k-anonymity works by keeping at least 'k' rows in an equivalence class, whereas the focus on l-diversity is having at least 'l' well represented values for each sensitive attribute in a quasi-identifier block. t-closeness makes it compulsory that the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold 't'. Differential privacy works by adding a noise to the results before publishing the data. However, it is a known fact that there is always a trade-off between privacy and accuracy. However, if we can train the model using the original data without compromising privacy, that model can overcome this accuracy-privacy trade-off. FL mechanisms can be employed here for training the global model without sharing the data to the central server, yet by using the original data at the clients' end. This will help in building a model with good accuracy, still without compromising privacy.

Prediction model formulation

The proposed AFib prediction model uses FL mechanism, Fed-CL to build a global model. This integrates CNN and LSTM to leverage their complementary strengths. The complexities associated with AFib diagnosis, especially in asymptomatic cases, demand early detection to prevent severe complications. The system can integrate wearable devices and artificial intelligence to process ECG signals promptly. The proposed methodology involves comprehensive data Pre-Processing, including transformation, normalization, and balancing techniques to address the challenges posed by variable-length and imbalanced datasets. The Synthetic Minority Over-sampling Technique (SMOTE) is used to address imbalance in the dataset. SMOTE produces synthetic samples for the minority class by interpolating the existing data which is available. It also increases the number of instances in the minority class, thereby balancing the dataset. SMOTE improves the model's learning of minority class decision boundaries by providing more balanced training data. This results to increased precision and recall for the minority class, improving the F1 score and model accuracy.

The FL model with hybrid CNN-LSTM is chosen for its efficacy in handling massive datasets, without compromising data privacy. The model training process, division of the dataset into train and test sets, and subsequent testing for performance measures are detailed. Figure 2 illustrates the general architecture of a FL model, showcasing its functionality in the AFib prediction context. The major steps in the proposed HAR prediction system using FL include model initialization, client selection, model distribution, local training, model aggregation, averaging and model re-distribution. The global model is initialized by the central server, which is then broadcasted to the chosen set of clients. This model is then used by the clients to train their local model. The model updates are then aggregated. The parameter updates from all the participating clients are then averaged. The global model uses the results to update the global model. The averaging process helps the global model to get benefited from the knowledge learned from various clients, yet preserving data privacy. Here, we use the FedAvg mechansim for parameter aggregation. The global model is again sent back to the participating clients and the process gets repeated until the model converges or until the desired performance is obtained.

The combination of CNN for local correlation feature extraction and LSTM for capturing front-to-back dependencies in ECG data is explained, providing a solid foundation for the subsequent discussions.

The problem of AFib prediction is mathematically formulated as depicted in equation 9, where, X, represent the set of features extracted from the ECG signals, each element x_i is a feature. Therefore, the goal is to predict the presence or absence of AFib, denoted as Y,based on the feature set X, as illustrated in the equation 9.

The equation 9 describes the mathematical approach to predicting AFib using ECG signals. The ECG traces the heart's electrical activity. This equation focuses on extracting specific characteristics (features) from that signal. These features, denoted by $\mathrm{x}_1, \mathrm{x}_2, \ldots \mathrm{x}_\mathrm{n}$ could represent various aspects like, Timing intervals between heartbeats, Wave shapes and amplitudes and Variability in heart rhythm. All the extracted features (x_1 to x_n) are grouped together as a set X. This set essentially captures the important characteristics of the ECG signal. The equation aims to predict the presence or absence of AFib based on the features in X.

$$\mathcal{X} = \{x_1, x_2, \ldots, x_n\} \quad \forall \mathcal{X} \in \{0, 1\} \tag{9}$$

The Fed-CL model for AFib prediction

Given the challenging nature of AFib diagnosis, the proposed framework aims to utilize the advancements in wearable devices and artificial intelligence for prediction. The framework encompasses ECG signal collection, pre-processing steps such as transformation, normalization, and balancing, and the application of a FL model

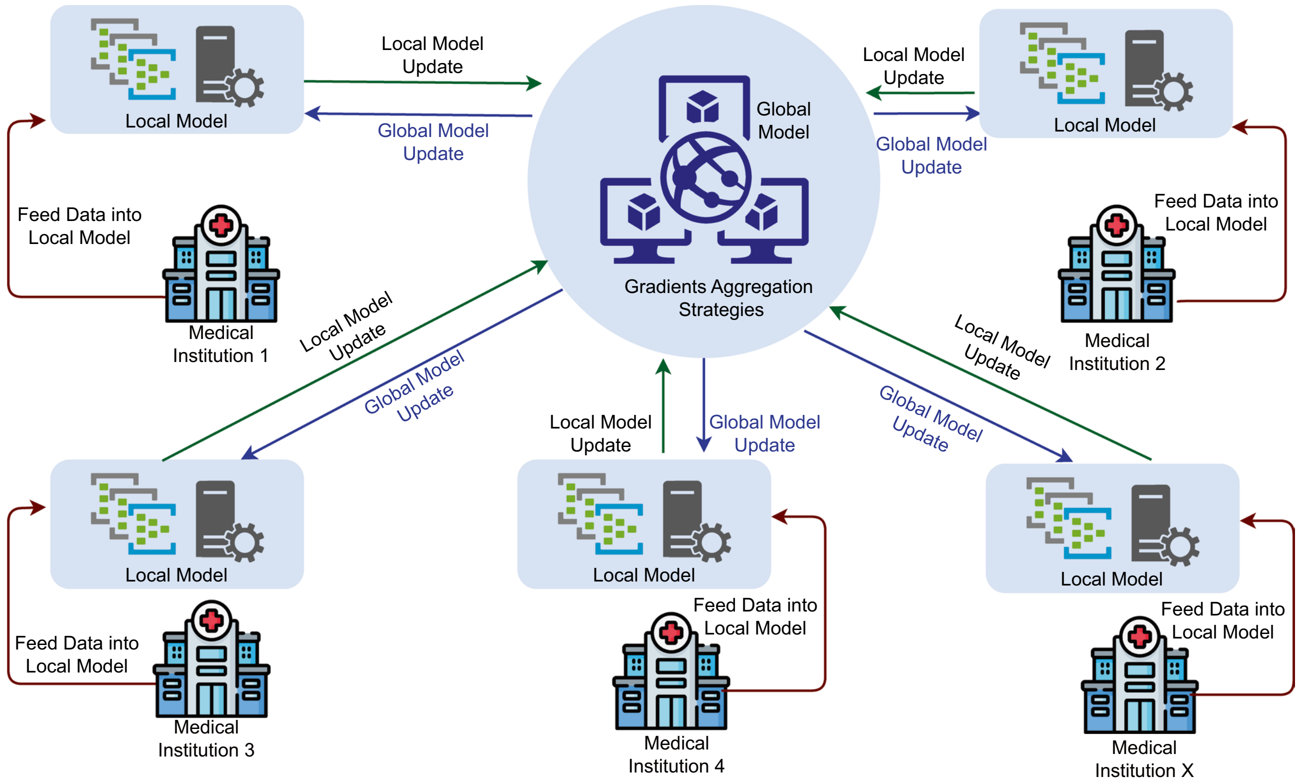


Figure 2. FL-based framework for atrial fibrillation prediction.

with hybrid CNN-LSTM model for classification. The unique architecture of the hybrid model is emphasized, as it significantly improves AFib prediction accuracy compared to traditional ML models. Notably, figure 3 illustrates the framework's architecture, underlining the key components and their interactions.

The discussions also stress the importance of early detection in improving treatment outcomes, enhancing the quality of life for patients, and reducing the economic burden associated with costly hospital interventions. The proposed methodology presents a novel approach by incorporating the ECG signals, addressing challenges associated with data velocity, and leveraging various mechanisms for improved accuracy. The proposed AFib prediction model, Fed-CL, integrates CNN and LSTM networks, denoted as MCNN-LSTM as in equation 10, to leverage their complementary strengths. CNN is used here for local correlation feature extraction and LSTM for capturing front-to-back dependencies in ECG data. The model is trained on a dataset D which is divided into training D_{train} and testing D_{test} sets.

$$\mathcal{MCNN} - \mathcal{LSTM} : X \to y \tag{10}$$

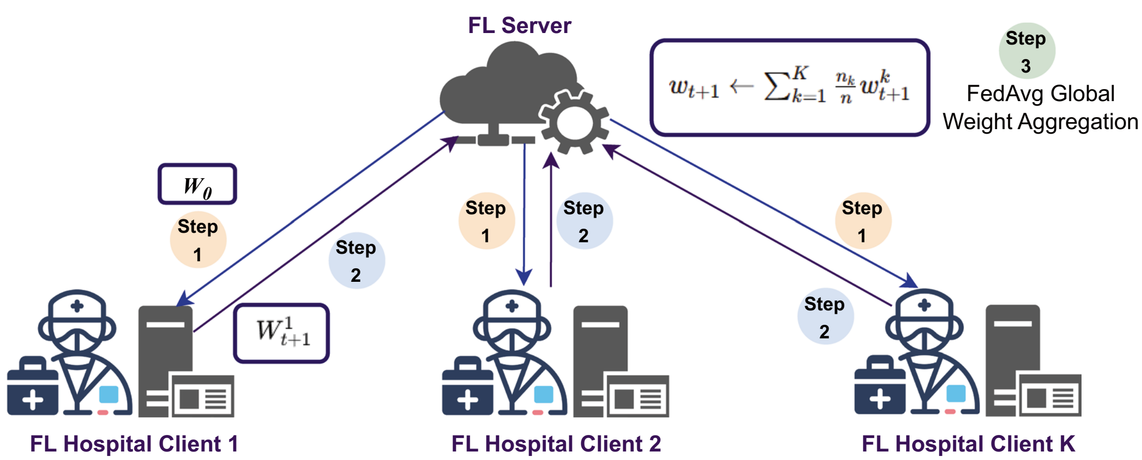


Figure 3. The Fed-CL Approach.

The nature of the system emphasizes the utilization of wearable devices and artificial intelligence to promptly process ECG signals. The problem involves comprehensive data pre-processing steps, including transformation, normalization, and balancing techniques. The pre-processing step results in a transformed feature set X. The transformation involves converting the variable-length and imbalanced datasets into a suitable format for model training. The Fed-CL model is chosen for its efficacy in handling massive datasets, which is crucial for timely predictions. The FedAvg algorithm uses weighted averaging, which weights each client's input to the global model based on the local dataset size. It ensures that clients with larger datasets, diverse datasets have a greater influence on the model, thus improving training stability and convergence across heterogeneous data.

The training process involves minimizing a loss function L over the training set D_{train} to optimize the model parameters as depicted in equation 11. The testing phase evaluates the performance of the model on the testing set D_{test}, providing measures such as accuracy, precision and recall as shown in equation 11. As illustrated in equations 11 and 12, θ represents the parameters of the MCNN-LSTM model.

$$\min_{\theta} \mathcal{L}\left(\mathrm{MCNN} - \mathrm{LSTM}\left(\mathcal{X}'_{\mathrm{train}}, \theta\right), \mathcal{Y}_{\mathrm{train}}\right) \tag{11}$$

$$Performance\ Measures = \mathrm{MCNN} - \mathrm{LSTM}\left(X'_{\mathrm{test}}, \theta\right) \tag{12}$$

The Fed-CL model employs MCNN-LSTM to train the global model. The algorithm for Fed-CL is provided in Algorithm 1. Initially, the global model is trained with the global weight w_{global} and is sent to the participating clients. Each client trains the local model with the data at the edge devices and updates the weights or associated parameters. The weights from various clients are then aggregated at the server and the global model is updated. The process repeats until the model converges or until the desired performance is attained. The equation for model aggregation using FedAvg algorithm is given in equation 13.

$$w_{t+1} = \Sigma_{c=1}^{C} \frac{n_k}{n} w_{t+1}^{i} \tag{13}$$

The Fed-CL model uses model compression and quantization before transferring the weights, thereby minimizing the size of the data sent during communication among the server and clients. Furthermore, Fed-CL model uses synchronization mechanism in which only significantly modified weights are shared. Increasing the number of local training epochs before sending updates to the server, effectively reduce communication overhead and enhance efficiency without compromising the frequency of updates necessary for model accuracy and convergence. Hence, the methodology incorporated in this study provides a robust foundation for the development and implementation of a AFib prediction system. Fed-CL, coupled with the hybrid model's architecture, showcases the potential of the proposed methodology in significantly impacting the early detection of AFib, ultimately leading to improved patient outcomes. The mathematical formulations, figure discussions, and detailed methodology contribute to the clarity and completeness of the proposed approach.

Server()
Initialize the global model weight w_{global} for atrial fibrillation prediction;
for *every communication round* $t = 1$ **to** T, **do**
    Select a client subset CS_t from the participating clients;
    for *every client* $i = 1$ **to** CS_t **do**
        Send the global weight w_{global} to the client 'i'.;
        Call the local atrial fibrillation prediction model, $w_{t+1}^{i} = clientlocal(i, w_t)$;
    end
    $w_{t+1} = \Sigma_{c=1}^{C} \frac{n_k}{n} w_{t+1}^{i}$;
    Assign the value of w_{t+1} to w_{global};
end

Clientlocal(i,w_loc**)**
for *each local epoch* $e = 1$ **to** E **do**
    for *each mini-batch* $b = 1$ **to** B **do**
        Compute the loss, $L(w_{loc}, b)$, where w_{loc} is the current local model weight.;
        Compute the gradient $\nabla L(w_{loc}, b)$. ;
        Update the weight as $w_t = w - \eta \nabla L(w_{loc}, b)$;
    end
end
Send the updated local weight, w_{loc} back to the server;

Algorithm 1. The Fed-CL algorithm for Atrial Fibrillation Prediction using FedAvg aggregation mechanism

Experimentation and analysis

The experimentation in this study revolves around the development and implementation of a predictive system for AFib utilizing ECG signals and FL techniques. As AFib often lacks prominent symptoms, early detection is crucial for effective treatment and prevention of complications such as stroke. The proposed methodology involves the collection of ECG signals from susceptible populations using wearable devices. The gathered data undergoes Pre-Processing, including transformation, normalization, and balancing, to ensure readiness for analysis.

Dataset

Atrial Fibrillation is usually diagnosed by electrocardiogram from the clinical diagnosis. The major parts in an ECG include the P wave, QRS complex and a T wave. The P wave indicates atrial depolarization and the Ventricular depolarization is represented by the QRS complex, that includes a Q wave, R wave, and S wave. Following the QRS complex, the T wave indicates ventricular repolarization. The sample ECG signals are shown in the Figure 4. The dataset description is shown in Table 2. In this work, we have segmented the ECG signals due to non-uniform distribution of the signals. As shown in Figure 5, the normal ECG has regular RR intervals and P wave is present, whereas, in the AFib signal the RR intervals are irregular and the P wave is absent.

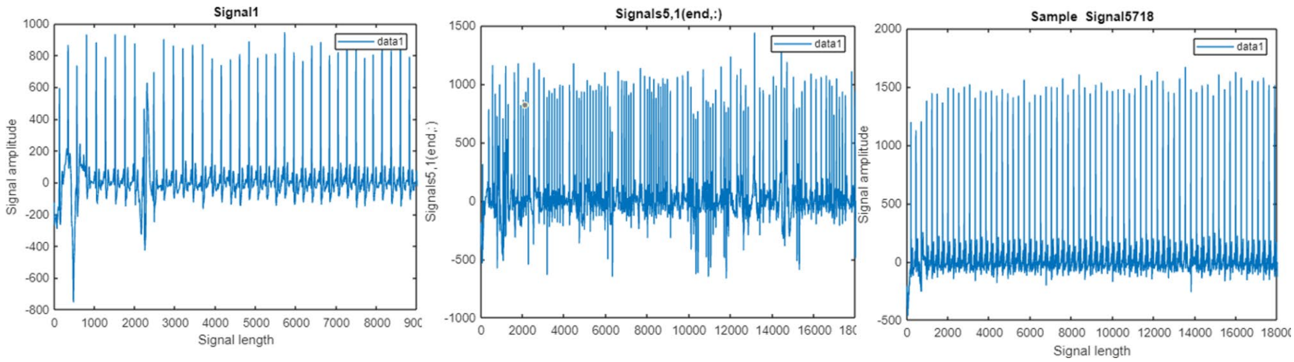


Figure 4. Sample ECG Signals.

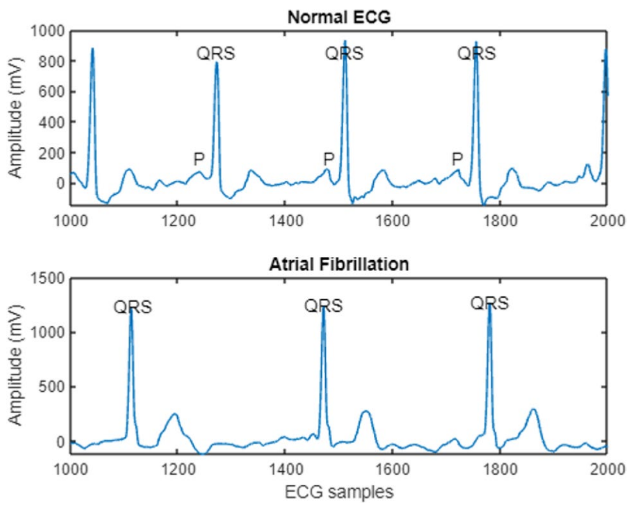


Figure 5. ECG Signals of Normal and AFib Conditions.

Parameter	Value
Dataset Size	5500 * 18500
Training Set	80%
Validation Set	10%
Test Set	10%
MAX Signal Length	8500
Sampling Frequency	300 HZ
Class Labels	2
Segmented Signal at Length	9000

Table 2. Dataset Description.

Dataset pre-processing

The ECG signals are susceptible to noises and over a period of time it loses its strength. In our work, as shown in the Figure 6., majority of the signals lost its strength between the signal length 9000 and 18000. Hence the the long duration of the ECG signals are segmented in to small segments of signal length not more than signal length = 9000. Table 3 depicts the time frequency patterns.

Since the signal data collected are of variable length and are highly imbalanced, the proposed system adopts various data pre-processing mechanisms such as data transformation, normalization, and balancing to make the data ready for further processing and analysis. The ECG signals obtained are converted into numerical data in the data transformation step. Min-max normalization is then applied to perform a linear transformation on the collected data. Furthermore, the SMOTE technique was applied to handle the data imbalance.

Classification and prediction

This includes data modeling, training, and testing the model and prediction phases. In order to handle massive datasets without compromising data privacy, the Fed-CL model with MCNN-LSTM was used. This model could yield better prediction results. CNN which has proved to be beneficial in image processing is employed here for extracting local correlation features and LSTM is used for capturing the front-to-back dependencies in the ECG data collected. Various machine learning models have been devised by researchers to detect and predict AFib. The entire dataset is divided into train and test sets. The train set is used to train the model. Once the training is done, the model is tested for various performance measures using the test set. Finally, the model can be used for predicting the output of new input data.

However, the two major challenges that exist are the velocity at which the data is collected and processed and the selection of the appropriate model to achieve better results. Our work is novel in the following aspects. The ECG signals collected are given as the input to the system. This method uses the heartbeat features of the input dataset to automatically predict the AFib in the provided signal. Since the signal data collected are of variable length and are highly imbalanced, the proposed system adopts various data pre-processing mechanisms such as data transformation, normalization, and balancing to make the data ready for further processing and analysis. Initially, a random forest algorithm is used to train the model and this proved to provide excellent results in terms of accuracy. However, as more data needs to be collected, processed, and analyzed, deep learning mechanisms can contribute more while dealing with big data sets. The hybrid CNN-LSTM approach can significantly improve the performance in AFib prediction by giving better accuracy, precision and recall when compared to the other approaches [18] [19] [20].

Even though the MCNN-LSTM model yielded better results in terms of various performance measures, preserving the privacy of data remains to be a significant concern, as the medical data is so sensitive and improper

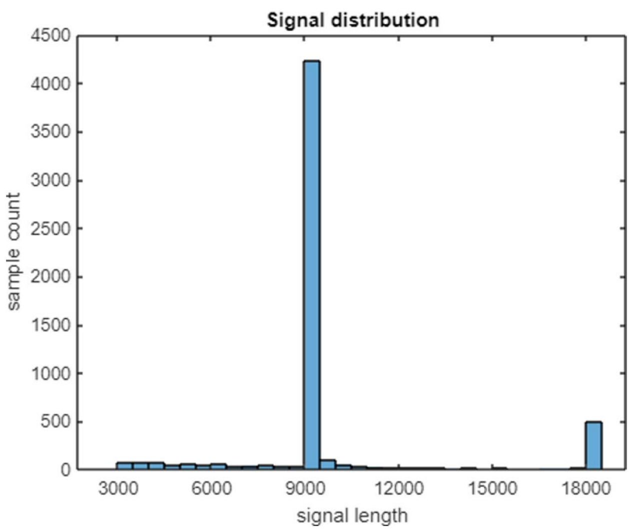


Figure 6. Distribution of the ECG signals.

Vector	Formula description
Mean (R)	Average of all R peeks
Mean (RR)	Average of all RR intervals
SD(R)	SD of all R peeks
SD(RR)	SD of all RR intervals

Table 3. Time Frequency Features.

handling may lead to various privacy risks. We employ Fed-CL model here that uses FL mechanism to train the model using the actual data, yet without sharing the same to the central server. The model uses the MCNN-LSTM architecture for model building.

We perform two experiments as detailed below. This is depicted in Figure 7.

Experiment 1: Training the classifier without feature extraction.

The occurrence of AFib in the ECG dataset is 12% of the whole dataset, which is minimum and hence it suffers from imbalanced dataset. The classifier would build a biased model which may simply classify all the signals as normal. This bias is termed as overfitting problem. Overfitting is a common challenge that is faced in binary classification. SMOTE is used to generate synthetic data samples for the AFib class. The CNN LSTM architecture is set with input size as 1, 150 hidden units, and the connected layer size is 2 as there are two classes. The prediction performance of Fed-CL is compared with other algorithms like Logistic regression, Decision Tree, Random Forest, Gradient Boosting, SVM, K-Nearest Neighbors, and Naïve Bayes.

Experiment 2: Training the classifier after feature extraction

The R wave in an ECG signal dataset is measured by finding the peak of the QRS complex. The QRS complex is the first major deflection in the ECG waveform and is caused by the depolarization of the ventricles. The R wave is the largest deflection in the QRS complex and is typically the easiest to identify. The Time-Frequency(TF) features like Mean (R), Mean (RR), standard deviation (R), standard deviation (RR) are extracted from the original ECG signals. Utilising the extracted features the various models are built as in experiment 1.

Results and discussion

The detailed evaluation of the machine learning models and Federated learning on the ECG data provides insights into the effects of feature extraction and SMOTE on model performance:

Analysis of accuracy score

Overall, MCNN-LSTM performs better than other models. When labels are included before feature extraction (FE) and combined with SMOTE (Synthetic Minority Oversampling Technique), a technique for addressing imbalanced class distributions, this model obtains the highest accuracy (96.32%) for classification. When compared to models without a feature extraction stage, the accuracy is typically higher when one is included (Before FE). Accuracy increases for most models when SMOTE is added to the data with feature extraction (After FE w/o SMOTE vs. After FE with SMOTE). This implies that correcting any possible class imbalances in the training data may help with the analysis. Fed-CL has competitive accuracy in the majority of settings, indicating that it is a useful tool for classification. While generally less accurate than MCNN-LSTM, DT (Decision Tree), RF (Random Forest), GB (Gradient Boosting), SVM (Support Vector Machine), and KNN (K-Nearest Neighbors) also demonstrate respectable accuracy. In all settings, Naive Bayes (NB) consistently has the lowest accuracy, suggesting that it may not be the best model for this particular classification analysis task. The results of the AFib prediction task using ECG signals are comprehensively analyzed and discussed based on the performance of various ML, DL, and FL models. Before and after feature extraction, with and without SMOTE, the models are evaluated for accuracy. Key observations include the MCNN-LSTM model showing a 96.32% accuracy, with feature extraction contributing positively. Fed-CL shows a perfect increase in accuracy after data balancing and feature extraction. Logistic Regression (LR) exhibits a modest 92.25%. DT and RF showcase substantial accuracy boosts post-feature extraction, reaching 94.21% and 92.87%, respectively. Gradient Boosting achieves 93.13% accuracy after feature extraction. SVM and KNN demonstrate varied impacts with SMOTE, while Naive Bayes struggles to capture ECG complexity. Logistic Regression stands out with remarkable improvement, emphasizing the role of tailored approaches for different models. The study provides valuable insights for selecting appropriate models and Pre-Processing techniques in clinical AFib prediction applications, acknowledging the interplay between feature engineering and oversampling. Figure 10 illustrates the overall accuracy of ML, DL,

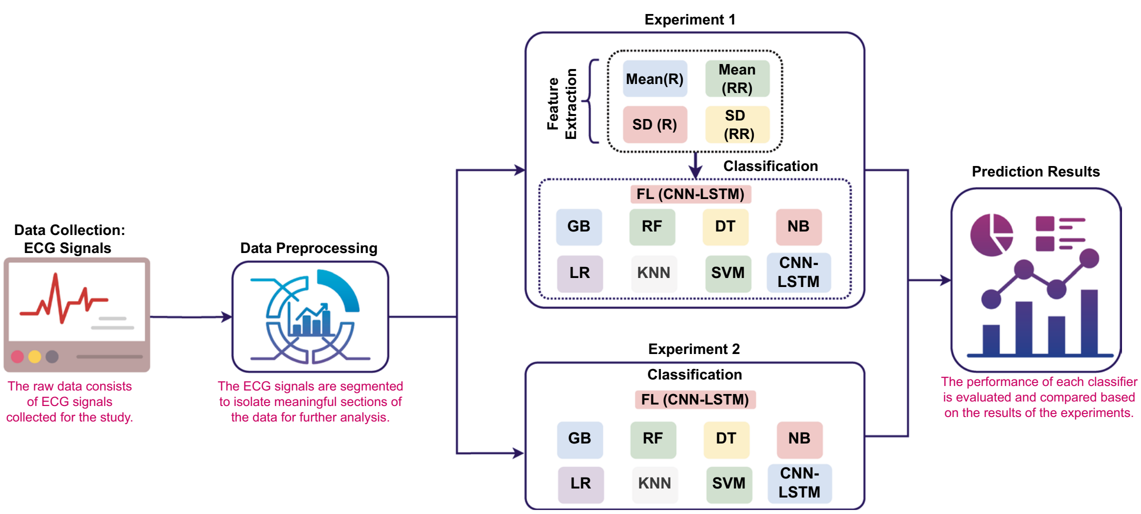


Figure 7. Framework of the study.

and FL models employed in this study. The accuracy and loss values of the server and four sample clients for the Fed-CL approach are depicted in Figures 8 and 9 respectively. The increase in accuracy values and decrease in loss values over a number of epochs demonstrate the efficiency of the Fed-CL model.

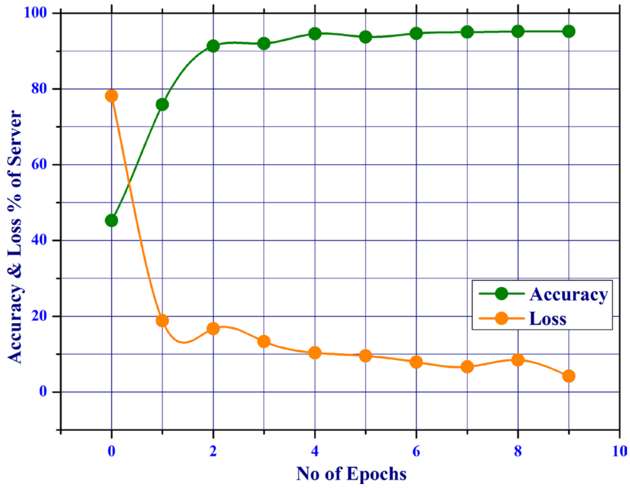
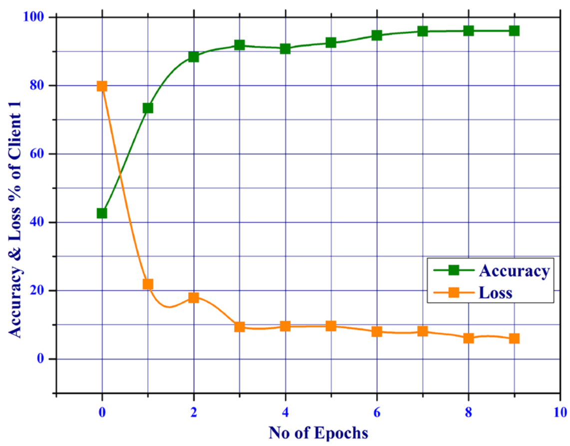
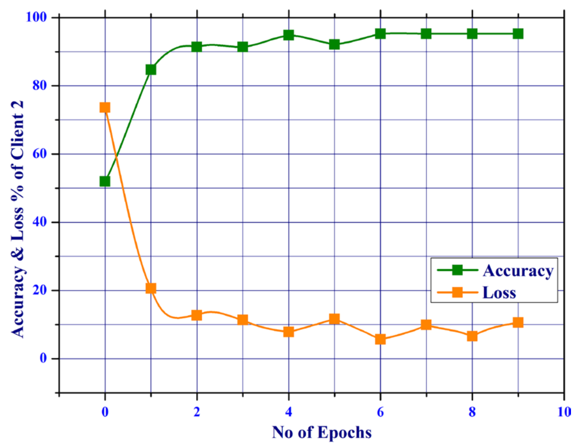


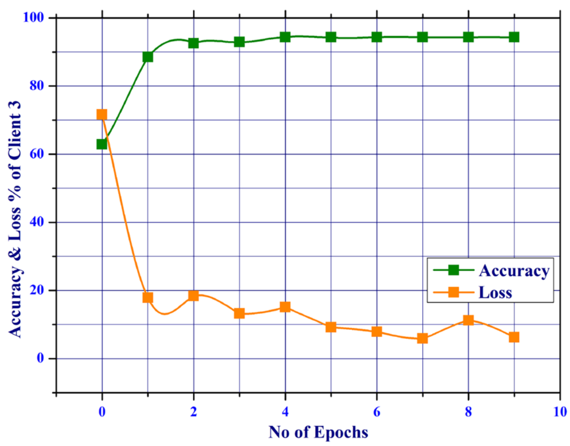
Figure 8. Training accuracy and loss- Server.



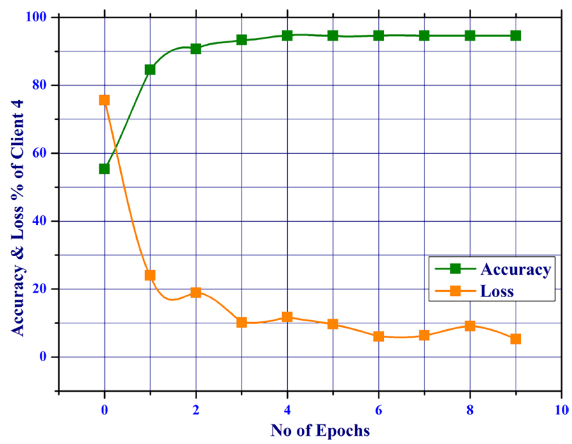
(a) Training accuracy and loss- Client 1.



(b) Training accuracy and loss- Client 2.



(c) Training accuracy and loss- Client 3.



(d) Training accuracy and loss- Client 4.

Figure 9. Accuracy and Loss- Sample Clients.

Analysis of precision score
The precision scores of diverse ML, DL and FL models in predicting AFib using ECG signals are thoroughly examined. Precision, crucial in binary classification, reflects a model's accuracy in identifying positive cases. The maximum precision is achieved by Fed-CL after FE and SMOTE (96.54%), indicating that a significant percentage of cases in the model classified as AFib are AFib. To prevent needless interventions based on false positives, this is essential. For MCNN-LSTM, precision increases from 90.35% to 95.73% after feature extraction with SMOTE, emphasizing the role of feature engineering and data balancing. LR sees a moderate increase to 90.45%. DT and Random Forest achieve remarkable precision (93.35% and 94.86%) after feature extraction, but SMOTE introduces a slight decline. GB shows perfect precision post-data balancing and feature extraction, with a minor decrease due to feature extraction. SVM decreases to 88.87% precision post-data balancing and feature extraction, while KNN and NB significantly improve from 89.62% to 93.75% and 88.36% to 94.69% precision post-data balancing and feature extraction. Fed-CL stands out, underlining the crucial role of feature extraction, while the impact of SMOTE varies across models. The findings stress the need for a nuanced approach considering both feature engineering and oversampling techniques for optimal AFib prediction in clinical applications. Figure 11 illustrates the overall precision of ML, DL, and FL models employed in this study.

Analysis of recall score
Recall measures how well the model was able to locate all of the real positive cases (people with AFib) in the dataset. The highest recall is typically achieved via Fed-CL. Fed-CL shows the highest Recall values across all configurations (Before/After Feature Extraction (FE) with/without SMOTE), demonstrating its efficacy in collecting the majority of real AFib cases. For the majority of models (apart from DT), adding a feature extraction step dramatically increases recall when compared to models that don't employ it (before FE). This implies that feature extraction aids in the models' acquisition of more discriminative features from the data, improving the models' ability to identify AFib cases. Adding SMOTE, a method to deal with unequal class distributions, can sometimes result in higher recall, especially for Fed-CL and MCNN-LSTM. This implies that there may be imbalanced classes in the dataset, possibly more examples of normal rhythms than AFib patients and that SMOTE aids in the models' ability to detect a higher number of actual AFib cases. Interestingly, even in the absence of feature extraction (Before FE w/o SMOTE), the DT model gets very high recall. This may be the result of the fact that DT models emphasize finding every possible scenario, even if doing so increases the possibility of some false positives, and rely less on complicated feature engineering. The Fed-CL combination with after FE and SMOTE, which achieves the greatest recall of 97.32%, is remarkable for its ability to accurately identify a significant percentage of real cases of AFib. The DT models exhibit good recall throughout, indicating that finding all possible occurrences of AFib is the primary objective. MCNN-LSTM with after FE and SMOTE (96.3% Recall) exhibits good performance, proving that it can handle imbalanced class management and feature extraction to capture the majority of AFib cases. In all settings, NB consistently has the lowest Recall, indicating that it may not be the best model to identify AFib instances in this dataset. Additionally, SVM models have lesser recall than some other models, which may indicate that they have trouble capturing specific AFib patterns. In situations involving medical diagnosis, where failing to identify true positives (undetected AFib) could have major repercussions, recall is essential. But it's crucial to think about striking a balance between precision (accurately recognizing positive cases) and recall. Due to false positives, a high recall rate combined with inadequate precision may result in needless interventions. Recall values can be affected by the particular dataset and the distribution of its classes. The model with the highest recall for AFib classification, particularly when feature extraction and SMOTE are used, is Fed-CL. But other models, such as DT, also seem promising in terms of giving priority to finding every possible case of AFib. The optimal model selection is contingent upon the particular clinical setting and the optimal trade-off between precision and recall. Figure 12 illustrates the overall recall of ML, DL, and FL models employed in this study.

Analysis of F1 score
To assess how well the ML, DL, and FL models perform overall in predicting AFib from ECG signals, the F1 score is analyzed. The F1-score, combining precision and recall, provides a balanced assessment of a model's effectiveness by considering both false positives and false negatives. In most cases, the addition of a feature extraction phase (After FE) improves the F1-score over the models without it (Before FE). This implies that the models are able to acquire more specific characteristics from the data, which in turn improves classification. The addition of SMOTE, improves the F1-score for a few models (Fed-CL, MCNN-LSTM, SVM). This shows that there may be imbalanced classes in the task (more examples of negative cases, in this study), and that SMOTE corrects this imbalance to improve model performance. MCNN-LSTM yields a competitive F1-score in the majority of configurations, indicating that it is a useful tool for classification analysis tasks. Though generally lower than Fed-CL, DT, RF, GB, and KNN all exhibit reasonable F1-score however, they could be further investigated with hyper-parameter tuning or alternative feature engineering techniques to potentially improve their performance. NB consistently has the lowest F1-score, indicating that Naive Bayes might not be the best model for this specific task. Based on the F1-score metric, Fed-CL with feature extraction and SMOTE is an effective sentiment analysis model overall. Figure 13 illustrates the overall accuracy of ML, DL and FL models employed in this study.

Performance analysis of ML, DL and FL models
This section provides an in-depth analysis of the performance of various ML, DL and FL models in predicting AFib based on ECG signals, considering key metrics such as Accuracy, Precision, Recall, and F1 Score. The analysis takes into account how various ML, DL and FL are affected by FE and SMOTE. Based on all four measures, the best-performing model is Fed-CL with feature extraction and SMOTE. Its high recall, accuracy, and

Accuracy percentage									
	Fed-CL	MCNN-LSTM	LR	DT	RF	GB	SVM	KNN	NB
Before FE w/o SMOTE	90.95	92.21	90.38	91.84	89.47	91.21	88.13	88.45	89.06
After FE w/o SMOTE	93.25	90.14	91.28	92.87	92.12	89.19	90.04	92.47	89.54
Before FE with SMOTE	92.04	91.32	89.85	92.12	90.84	88.54	90.21	89.81	91.83
After FE with SMOTE	95.25	96.32	92.25	94.21	92.87	93.13	91.45	93.84	95.05
Precision Percentage									
	Fed-CL	MCNN-LSTM	LR	DT	RF	GB	SVM	KNN	NB
Before FE w/o SMOTE	90.91	90.35	87.86	92.71	91.71	88.08	91.17	89.62	88.36
After FE w/o SMOTE	93.7	92.03	87.68	92.77	93.33	87.11	92.74	92.15	91.85
Before FE with SMOTE	91.04	90.95	90.06	92.6	92.25	89.03	89.84	88.5	89.16
After FE with SMOTE	96.54	95.73	90.45	93.35	94.86	92.6	88.87	93.75	94.69
Recall Percentage									
	Fed-CL	MCNN-LSTM	LR	DT	RF	GB	SVM	KNN	NB
Before FE w/o SMOTE	89.61	81.23	94.19	97.93	94.56	66.14	89.99	81.69	75.42
After FE w/o SMOTE	95.14	85.52	96.82	96.88	94.31	87.73	95.3	95.33	78.45
Before FE with SMOTE	90.98	92.81	86.44	88.59	92.09	86.03	90.35	87.64	90.58
After FE with SMOTE	97.32	96.3	93.17	93.96	94.74	88.59	92.32	93.42	96.8
F1 Score Percentage									
	Fed-CL	MCNN-LSTM	LR	DT	RF	GB	SVM	KNN	NB
Before FE w/o SMOTE	90.02	85.55	90.91	95.25	93.11	75.55	90.38	85.47	81.38
After FE w/o SMOTE	94.41	88.66	92.02	94.78	93.82	87.42	94.01	93.71	84.62
Before FE with SMOTE	91.01	91.87	88.21	90.55	92.17	87.51	90.1	88.07	89.86
After FE with SMOTE	96.93	96.01	91.79	93.65	94.8	90.55	90.56	93.59	95.73

Table 4. Comparative analysis.

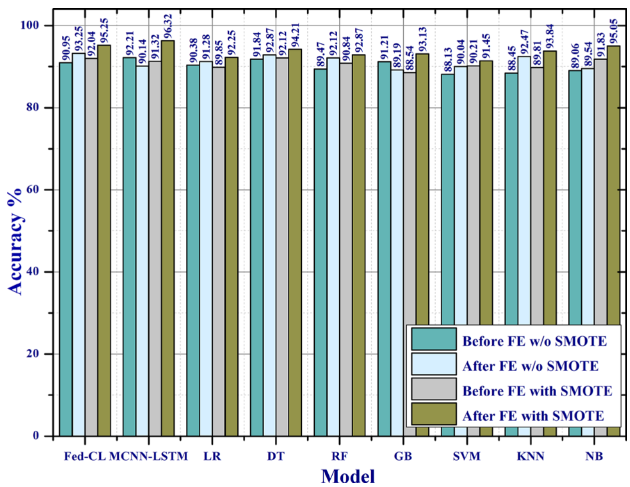


Figure 10. Analysis of Accuracy Score.

precision show that it is a useful tool for accurately categorizing AFib instances with the fewest possible mistakes. As for prioritizing high recall, which could be preferable in this case, the Decision Tree approach appears promising. Fed-CL, constantly performs best in the majority of settings, particularly when combined with SMOTE and feature extraction. It provides an excellent mix between recall (recording the majority of AFib cases) and precision (identifying AFib properly). MCNN-LSTM performs competitively throughout the board, especially when it comes to SMOTE and feature extraction. It offers a potent substitute for Fed-CL. DTs may have inferior precision even while obtaining good recall without Feature Extraction. Even with some false positives, they could be appropriate if identifying all possible AFib instances is of utmost importance. The efficacy varies with other Models (LR, RF, GB, SVM, KNN, NB). Adding a feature extraction phase typically results in a significant improvement over models without it in all measures (Before FE). This emphasizes the significance of the use of feature extraction in converting unprocessed data into more discriminating features necessary for precise AFib classification. Certain measurements have fundamental trade-offs. For example, DT prioritize recognizing all

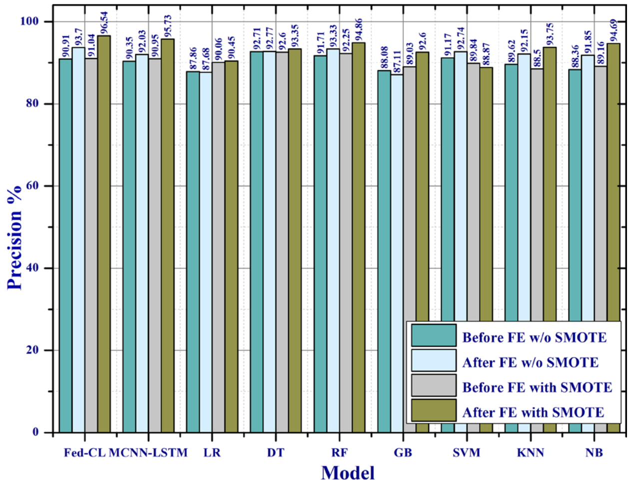


Figure 11. Analysis of Precision Score.

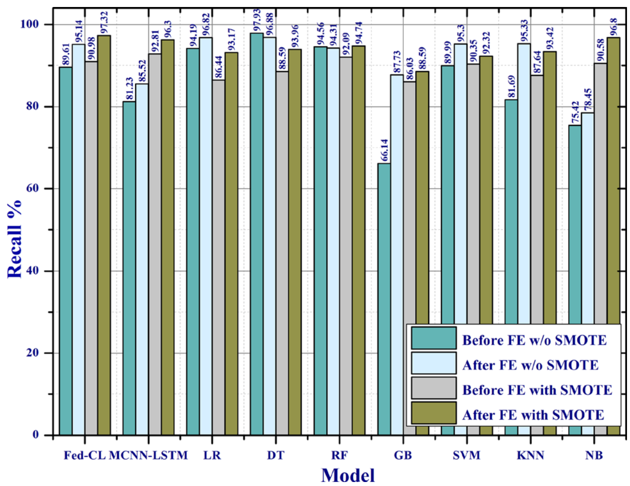


Figure 12. Analysis of Recall Score.

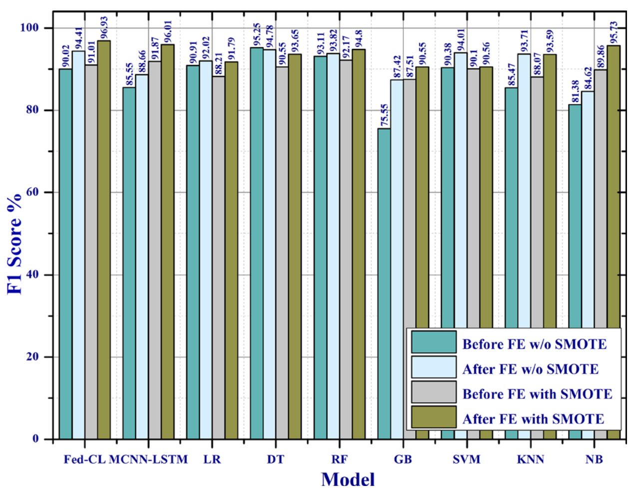


Figure 13. Analysis of F1 Score.

possible AFib situations, as evidenced by their high recall (97.93%) even in the absence of feature extraction (Before FE w/o SMOTE). Their precision (92.71%), however, may be lower, which implies that some false positives may be included. Table 4 provides an overview of the results of various performance mesaures of various ML, DL and FL algorithms.

Impact of SMOTE

SMOTE appears to significantly improve the performance of different machine learning models for atrial fibrillation prediction, according to the result table. SMOTE improved accuracy for all models (Fed-CL, MCNN-LSTM, etc.) and feature extraction scenarios (Before vs. After). The range of improvement is 2.21% to 7.21%. SMOTE improved precision for most models in a manner similar to accuracy, with gains ranging from 1.69% to 6.54%. Recall is more unevenly affected by SMOTE. Certain models, such as Fed-CL and DT, demonstrated a noteworthy enhancement of up to 15.69% in recall, but SVM demonstrated a decline in recall. For several models (such as Fed-CL, RF), SMOTE had a favorable effect on the F1 score, which is consistent with the trend shown in recall.

This section concludes by providing a comprehensive summary of the results obtained. The study compared the effectiveness of different machine learning models for atrial fibrillation (AFib) prediction, and the findings are summarized in the table. The advantages of feature extraction are demonstrated by a straight comparison of performance metrics before and after this process. The accuracy of the Fed-CL, MCNN-LSTM, KNN and NB increased dramatically approximately 4% after to feature extraction following extraction SMOTE. This notable improvement emphasizes how important feature extraction is to improving model performance. Four measures are used to assess the models: F1 score, Accuracy, Precision, and Recall. Using FE generally results in better model performance than without using FE for all measures (Accuracy, Precision, Recall, and F1-score). In certain situations, the improvement can be significant, ranging from a few percentage points to more than 10 percentage points. Depending on the model and metric, SMOTE seems to improve the performance of most models. However, this effect varies. For the majority of models, accuracy rises with SMOTE, typically by 1.41% to 5.25%. SMOTE increases precision for most models in a manner similar to accuracy, with gains ranging from 0.32% to 6.54%. SMOTE has a more uneven effect on recollection. Some models exhibit a reduction or no change, while others demonstrate a notable improvement (up to 15.69%). The pattern for F1 score is comparable to that of recall, showing a positive influence on certain models and a negative one on others.Fed-CL with SMOTE and Feature Extraction performs best overall, with accuracy higher than 96%, in all criteria.

Conclusion

The culmination of this study brings forth significant insights and advancements in the domain of AFib detection using ECG signals and employing various ML, DL and FL techniques. The research has successfully addressed the challenges associated with the asymptomatic nature of AFib, emphasizing the importance of early detection for timely intervention and prevention of severe complications. Thus, the proposed predictive system, employing the Fed-CL model, showcases promising results in terms of accuracy, precision and recall. The experimentation phase, involving extensive data pre-processing and model training, validates the efficacy of the system in handling the ECG signals from susceptible populations. The utilization of wearables and advanced artificial intelligence mechanisms provides a robust foundation for enhancing AFib detection and subsequently improving patient outcomes. The significance of early AFib detection is highlighted by its potential to not only ensure successful treatment and possible cure, but also to enhance the overall quality of life. The study emphasizes the role of public health programs and periodic screening examinations, supported by the proposed system, in detecting unnoticed diseases and preventing their progression. Furthermore, the early diagnosis facilitated by the system can lead to reduced reliance on costly hospital interventions, thereby contributing to more efficient healthcare resource utilization.

The study sets the stage for several avenues of future research and development. Firstly, continuous refinement and optimization of the proposed predictive system will be essential to accommodate evolving technological landscapes and enhance its adaptability to diverse populations. Additionally, further exploration of different FL architectures, feature extraction techniques, and data balancing methods could contribute to even more accurate and robust AFib detection models. The integration of additional physiological parameters and health-related data into the predictive system could broaden its scope and enhance its predictive capabilities. Collaboration with healthcare professionals and institutions will be crucial for real-world validation and the incorporation of the system into existing healthcare practices. The study envisions the integration of the proposed system into wearable devices, making it readily accessible to individuals at risk of AFib. This move towards user-friendly applications can facilitate widespread adoption and contribute to community-based health monitoring.

In conclusion, the research not only advances the field of cardiovascular health by introducing an effective AFib detection system, but also paves the way for future innovations. The continuous evolution and refinement of such systems hold the potential to revolutionize preventive healthcare practices, ultimately contributing to improved patient care and outcomes in the domain of cardiac arrhythmia detection and management.

Data availibility

The dataset generated and analysed during the current study are available in the PhysioNet repository, WEB LINK TO DATASETS: https://physionet.org/content/challenge-2017/1.0.0/.

Received: 8 April 2024; Accepted: 26 August 2024
Published online: 09 September 2024

References

1. Freedman, B. *et al.* World heart federation roadmap on atrial fibrillation–a 2020 update. *Global heart* **16** (2021).
2. Patel, P. A., Ali, N., Hogarth, A. & Tayebjee, M. H. Management strategies for atrial fibrillation. *J. Royal Soc. Med.* **110**, 13–22 (2017).
3. Ponamgi, S. P. *et al.* Screening and management of atrial fibrillation in primary care. *Bmj* **373** (2021).
4. Chyou, J. Y. *et al.* Atrial fibrillation occurring during acute hospitalization: A scientific statement from the american heart association. *Circulation* **147**, e676–e698 (2023).
5. Lippi, G., Sanchis-Gomar, F. & Cervellin, G. Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. *Int. J. Stroke* **16**, 217–221 (2021).
6. Anter, E., Jessup, M. & Callans, D. J. Atrial fibrillation and heart failure: treatment considerations for a dual epidemic. *Circulation* **119**, 2516–2525 (2009).
7. Spartera, M. *et al.* The impact of atrial fibrillation and stroke risk factors on left atrial blood flow characteristics. *European Heart J.-Cardiovasc. Imaging* **23**, 115–123 (2022).
8. Qiu, D., Peng, L., Ghista, D. N. & Wong, K. K. Left atrial remodeling mechanisms associated with atrial fibrillation. *Cardiovasc. Eng. Technol.* **12**, 361–372 (2021).
9. Rizwan, A. *et al.* A review on the state of the art in atrial fibrillation detection enabled by machine learning. *IEEE Rev. Biomed. Eng.* **14**, 219–239 (2020).
10. Pereira, T. *et al.* Photoplethysmography based atrial fibrillation detection: a review. *NPJ Digital Medicine* **3**, 1–12 (2020).
11. Quer, G., Arnaout, R., Henne, M. & Arnaout, R. Machine learning and the future of cardiovascular care: Jacc state-of-the-art review. *J. Am. College Cardiol.* **77**, 300–313 (2021).
12. Krittanawong, C. *et al.* Deep learning for cardiovascular medicine: a practical primer. *Eur. Heart J.* **40**, 2058–2073 (2019).
13. Abd-Ellah, M. K., Khalaf, A. A., Gharieb, R. R. & Hassanin, D. A. Automatic diagnosis of common carotid artery disease using different machine learning techniques. *J. Ambient Intell. Human. Comput.* **14**, 113–129 (2023).
14. Alzeer, A. H. *et al.* Using machine learning to reduce unnecessary rehospitalization of cardiovascular patients in saudi arabia. *Int. J. Med. Inform.* **154**, 104565 (2021).
15. Sadad, T. *et al.* Detection of cardiovascular disease based on ppg signals using machine learning with cloud computing. *Computational Intelligence and Neuroscience* **2022** (2022).
16. Wegner, F. K. *et al.* Machine learning in the detection and management of atrial fibrillation. *Clin. Res. Cardiol.* **111**, 1010–1017 (2022).
17. Siontis, K. C., Yao, X., Pirruccello, J. P., Philippakis, A. A. & Noseworthy, P. A. How will machine learning inform the clinical care of atrial fibrillation?. *Circulation Res.* **127**, 155–169 (2020).
18. Alhusseini, M. I. *et al.* Machine learning to classify intracardiac electrical patterns during atrial fibrillation: machine learning of atrial fibrillation. *Circulat. Arrhythmia Electrophysiol.* **13**, e008160 (2020).
19. Baj, G. *et al.* Comparison of discrimination and calibration performance of ecg-based machine learning models for prediction of new-onset atrial fibrillation. *BMC Med. Res. Methodol.* **23**, 1–10 (2023).
20. Bhardwaj, A. *et al.* A holistic overview of artificial intelligence in detection, classification and prediction of atrial fibrillation using electrocardiogram: a systematic review and meta-analysis. *Arch. Comput. Methods Eng.* **30**(7), 4063–79 (2023).
21. Morvari, A. & Moghaddam, R. K. Detecting atrial fibrillation from ecg signal using hybrid convolutional neural network with ant-lion optimization. *Multimed. Tools Appl.* **83**, 1–13 (2023).
22. Xie, J., Stavrakis, S. & Yao, B. Automated identication of atrial fibrillation from single-lead ecgs using multi-branching resnet. arXiv preprint arXiv:2306.15096 (2023).
23. Jiang, Y., Zheng, P. & Lai, D. A semi-supervised algorithm for atrial fibrillation attack prediction using convolution auto-encoder of time series signal. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1–4 (IEEE, 2023).
24. Duranta, D. *et al.* Enhancing atrial fibrillation detection accuracy: A wavelet transform filtered single lead ecg signal analysis with artificial neural networks and novel feature extraction. *Machine Learn. Appl.* **12**, 100472. https://doi.org/10.1016/j.mlwa.2023.100472 (2023).
25. Aldughayfiq, B., Ashfaq, F., Jhanjhi, N. & Humayun, M. A deep learning approach for atrial fibrillation classification using multi-feature time series data from ecg and ppg. *Diagnostics* **13**, 2442 (2023).
26. Jimenez Gutierrez, D. M., Hassan, H. M., Landi, L., Vitaletti, A. & Chatzigiannakis, I. Application of federated learning techniques for arrhythmia classification using 12-lead ecg signals. In *International Symposium on Algorithmic Aspects of Cloud Computing*, 38–65 (Springer, 2023).
27. Yaqoob, M. M. *et al.* Modified artificial bee colony based feature optimized federated learning for heart disease diagnosis in healthcare. *Appl. Sci.* **12**, 12080 (2022).
28. Yaqoob, M. M., Nazir, M., Khan, M. A., Qureshi, S. & Al-Rasheed, A. Hybrid classifier-based federated learning in health service providers for cardiovascular disease prediction. *Appl. Sci.* **13**, 1911 (2023).
29. Asif, R. N. *et al.* Detecting electrocardiogram arrhythmia empowered with weighted federated learning. *IEEE Access* (2023).
30. Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M. & Baik, S. W. Action recognition in video sequences using deep bi-directional lstm with cnn features. *IEEE access* **6**, 1155–1166 (2017).
31. Alzubaidi, L. *et al.* Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *J. Big Data* **8**, 1–74 (2021).
32. Sweeney, L. k-anonymity: A model for protecting privacy. *Int. J. Uncertainty, fuzziness knowledge-based syst.* **10**, 557–570 (2002).
33. Machanavajjhala, A., Kifer, D., Gehrke, J. & Venkitasubramaniam, M. l-diversity: Privacy beyond k-anonymity. *Acm transactions on knowledge discovery from data (tkdd)* **1**, 3 (2007).
34. Li, N., Li, T. & Venkatasubramanian, S. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd international conference on data engineering*, 106–115 (IEEE, 2006).
35. Dwork, C. Differential privacy. In *International colloquium on automata, languages, and programming*, 1–12 (Springer, 2006).

Acknowledgements

This research has been funded by Deputy for Research & Innovation, Ministry of Education through Initiative of Institutional Funding at University of Ha'il - Saudi Arabia through project number IFP-22 126.

Author contributions

All listed authors have substantially contributed to the manuscript and have approved the final submitted version, the description of each author's specific work and contributions are as follows: The contribution statement: The authors confirm contribution to the paper as follows: study conception and design: Fayez Saud Alreshidi, Rajeswari Chengoden. data collection: Fayez Saud Alreshidi, Mohammad Alsaffar, Naif Khalaf Alshammari; analysis and interpretation of results: Fayez Saud Alreshidi, Mohammad Alsaffar, Rajeswari Chengoden, Naif

Khalaf Alshammai; draft manuscript preparation: Fayez Saud Alreshidi, Mohammad Alsaffar, Rajeswari Chengoden, Naif Khalaf Alshammari.

Competing interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Additional information

Correspondence and requests for materials should be addressed to R.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.